"I Call Shotgun!": An Evaluation of Mixed-Initiative Control for Novice Users of a Search and Rescue Robot*

David J. Bruemmer, Ronald L. Boring, Douglas A. Few, Julie L. Marble, Miles C. Walton
Human, Robotic, and Remote Systems Department
Idaho National Engineering and Environmental Laboratory
Idaho Falls, ID, U.S.A.
{bruedj, borirl, fewda, marbjl, mwalton}@inel.gov

Abstract - *The quality of human-robot interaction trails other advances in robotics and may prove to be a limiting factor when deploying remote, mobile robots for critical applications. One reason is that most autonomous robot behaviors are not robust and often degrade in unstructured environments. Another reason is that the design of human-robot interaction (HRI) and interfaces fails to follow basic usability principles or be informed by basic concepts of human-computer interaction. To address both these challenges, we have used a development cycle of iterative usability testing and redesign to hone both our interface and the robot behaviors that support it. The present paper presents results from a wide swathe of over 100 novices who used the resulting system to accomplish a real-world search and detection task. The current interface proved to be highly usable by novices, regardless of age or gender. The study demonstrates the utility of effective robot autonomy and examines the benefits of mixed-initiative control. In particular, the study compares the performance achieved when the robot takes initiative to support human driving versus the case where the human takes initiative to support autonomous robot driving. Results indicate that performance is better when the robot is in the driver's seat. Optimal performance was achieved when the operator focuses on the search and rescue task and provides only intermittent direction to the robot.*

Keywords: Human-robot interaction, usability evaluation, novice performance, urban search and rescue.

1 Introduction

Remote robotic operations often involve operator workload constraints, limited communication bandwidth, long distances, and/or a lack of visible environmental features. Given these constraints, teleoperation seems a poor choice and yet almost all mobile ground systems in use today place the human squarely in the driver's seat. Why? Despite the recognized need for autonomy, the performance of most intelligent robots does not approach that of a human operator and often fails to support operator trust. Unless we improve both the robustness of robot behavior and the flexibility of human-robot interaction methods, robots will continue to be excluded from the many environments and tasks where teleoperation is unsuitable. To address this challenge, we believe that rigorous, real-world human-robot interaction (HRI) evaluations must lead the way.

Work in this area has already begun. There are taxonomies for effective HRI [7] and metrics for evaluating HRI [6], but these have not yet translated into effective HRI evaluations in practice. Typical HRI evaluations utilize an informal testing scenario and still employ robotic designers as test participants. Such experiments inaccurately predict the performance to be expected of actual operators. Additionally, many HRI evaluations occur with simulated robots and therefore say little about operator trust and the robustness of real-world robotic behavior.

Yanco, Drury, and Scholtz [8] have identified two major shortcomings in current HRI evaluations. First, HRI evaluations typically fail to test the intended user of the system. Often, the designers of the system are also the test users. Such evaluation is flawed, because system designers possess a much higher system understanding and proficiency than do the actual users of that system. Thus, the designers represent an upper bound of expected performance, and evaluations may fail to identify the difficulties that an actual HRI user might experience. The second shortcoming is that HRI evaluations are commonly informal, precluding careful empirical control. As a consequence, most HRI evaluations fail to provide objective or conclusive results. Yanco et al. [8] do not dismiss offhand the value of current HRI evaluation methods. Rather, Yanco et al. aim to complement existing HRI evaluation methods by pointing the way toward more effective evaluation.

In response to the shortcomings identified in [8], we employed a carefully controlled experimental setting (i.e., an urban search and rescue arena) with objective performance metrics (i.e., the number of objects located and identified in the arena in a fixed amount of time). As suggested by Yanco et al. [8], we avoided evaluating the interface with system designers or seasoned operators. Instead, we enlisted novice users of robotic systems in our evaluation. Novice users do not directly represent the target user for a search and rescue robotic system. Most likely, search and rescue personnel will have been given a significant level of training.

We are not primarily interested in Search and Rescue, but rather in the general question of how humans and robots can best cooperate. Our applications include countermine operations, remote characterization of high radiation environments and military reconnaissance. However, we have subjected our system to HRI studies with Search and Rescue experts from the Federal Emergency Management Agency (FEMA) and found that just because a participant has expertise in the problem of search and rescue in general does not mean that they are necessarily good at operating robots (although some clearly are). In fact, as robots become more capable and intelligent, it may be that the protocols and methods utilized by Search and Rescue personnel change drastically.

We believe that by opting for novice users, we maximized both the relevance of our study to multiple applications and our evaluation's sensitivity to interface shortcomings. Such an evaluation does not, of course, preclude the necessity of further evaluation with the actual target users. An evaluation of novice users does, nonetheless, provide a baseline performance measure using a greater number of participants than would otherwise be possible.

This is not the first human participant test of our HRI and control architecture. A key, underlying component to successful, usable design is iterative testing and redesign based on the results of this testing. Previously, novice and expert users were asked to search a building in a pseudo search and rescue task, making use of the different levels of our variable control architecture [5]. This experiment had several interesting results, among them that users who had expertise driving teleoperated robotic systems were less trusting and less willing to use higher levels of system autonomy. This preliminary study served as impetus for us to change several aspects of the control architecture including the video transmission, the frequency and duration of robot initiative and the robot's sensitivity to obstacles. In addition, this first experiment prompted us to change the interface to allow online configuration of the robot's sensing and initiative. We also included representations that indicate the robot's confidence in its own sensors and decisions in order to support appropriate levels of operator trust. We believe that through this iterative design process, a more robust system was created than could have been achieved from informal testing or simulation-based studies.

2 Robotic Search and Rescue

Urban search and rescue has become a major implementation area for robotics, because robots can be sent into environments that may be unsafe for human search personnel [3]. Urban search and rescue is a true test bed for robots, in that it tests the robustness of the robotic hardware and the overall agility of the robot.

Urban search and rescue also tests the HRI. Behind the physical agility of the robot is the ability of the operator to control the robot's movement and actions in an effective way. A usable interface can facilitate successful search and rescue operations, potentially saving lives. Conversely, an interface that is not usable can hamper search and rescue operations, ultimately risking lives. Despite the obvious importance of good HRI for urban search and rescue, there is surprisingly little consistency across the interfaces that operators use to control robots [8].

With human lives at stake, it is important to test search and rescue robots before they are put to use in an actual life and death situation. The National Institute of Standards (NIST) has developed test arenas for urban search and rescue, which have been used in robotics competitions such as RoboCup Rescue and the annual conference of the American Association for Artifical Intelligence (AAAI). The NIST test arenas are classified into three color-demarked categories [4]. The yellow category is the easiest urban-type arena, with minimal obstructions and clear visibility. The orange category encompasses increased complexity in moving through the environment and locating objects of interest. The orange arena is spread over two physical levels and may contain obstacles such as stairs, requiring greater robot agility. The red category is the most difficult, featuring highly obstructed terrain and mostly buried objects as would be typical in the rubble aftermath of a collapsed building. In our study, we approximated a NIST category yellow urban search and rescue arena.

3 Study

3.1 Participants

The present study included 107 participants drawn at random from attendees of Idaho National Engineering and Environmental Laboratory's (INEEL's) annual community exposition. The participants consisted of 46 females and 61 males, ranging in age from 3 to 78 years old, with a mean age of 14. It could be argued that attendees of a science and engineering exposition are likely to be more technologically savvy than the general populous. However, when questioned, none of the participants had experience in remote system operations, thus qualifying them as novice users of robotic interfaces.

3.2 Robot Description

The robot used in the present study was a wheeled ATVRjr manufactured by iRobot (see Figure 1), which measures approximately 67 cm long x 54 cm wide including the wheelspan. The ATVRjr was augmented by a sensor array designed by the INEEL [1] and includes infrared sensors, bump sensors, scanning laser, ultrasonic sensors, visual camera, thermal camera, tilt sensor, and gyro. The sensory information is used by the robot itself during autonomous operation and is also available in the form of meaningful abstractions to the robot operator. The robot provides a video feed to the operator from a forward mounted camera. The video signal and the sensory data are fed to the control station via a wireless link.

The robot platform was modified for use in hazardous environments for tasks such as radioactivity monitoring at decommissioned nuclear sites [2]. Compared to conventional monitoring by humans, the robot can accomplish a full environmental analysis with minimal human exposure to radioactivity. The sites, consisting of vacant buildings once used for handling radioactive materials, contain terrain and obstacles very similar to an urban search and rescue setting. The robot's robustness and agility in these harsh environments translate into equivalent search and rescue strengths. However, the robot is not directly equipped for rescue operations. As a search and rescue robot, its primary goal is to map dangers in the environment and pinpoint the location of objects such as survivors. The robot performs searches for subsequent human rescue operations.

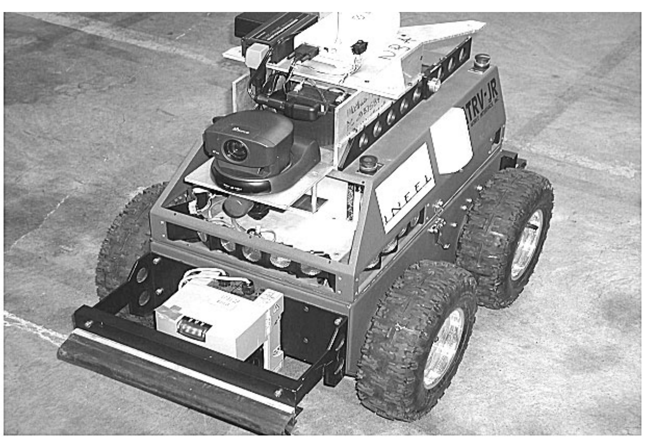

Figure 1. INEEL's augmented ATVRjr robot

3.3 Interface Description

The robotic interface is the culmination of iterative usability testing and redesign [5]. In designing the interface, we attempted to strike a balance between ease of robotic control and the rich information display necessary for monitoring hazardous environments or conducting search and rescue. The interface consists of a single touchscreen display containing five sizeable windows (see Figure 2). The upper left-hand window on the screen contains a video feed from the robot as well as controls for adjusting the camera. The upper right-hand window contains status indicators and controls for the robot's sensors. The lower right-hand window features movement status indicators and controls as well as a mode selector for different levels of robot autonomy. The lower central window provides an emerging map of the environment as determined by a simultaneous mapping and localization algorithm developed by the Naval Research Laboratory. Finally, The lower left-hand window contains information about the robot's operational status.

Many features of the interface are geared toward expert operators. For example, the interface offers two windows with primarily status indicators to reflect the robot's operational state and input from the advanced sensor array. Despite these advanced features, it is easy for the novice user to focus on two diagonally situated windows. The upper left-hand window contains the video display, which allows the user to see from the robot's vantage point. The lower right-hand window contains the main movement controls for the robot, whereby directional arrow buttons actuate the movement.

Control of the robot can be achieved by touching appropriate areas of the display. The effect of these touches depends on the mode of autonomy. When in direct control, operators primarily give directional commands using the joystick. The participants were explicitly instructed on using the onscreen controls as well as using a joystick to control the robot's movement.

Four robot initiative modes are available in the interface [1,2], affording the robot different types of behavior and levels of autonomy. *Tele Mode* is a fully manual mode of operation, in which the operator must manually control all robot movement. *Safe Mode* is similar to Tele Mode, in that robot movement is dependent on manual control. However, in Safe Mode, the robot is equipped with a level of initiative that prevents the operator from colliding with obstacles. In *Shared Mode*, the robot relieves the operator from the burden of continuous navigation by finding the optimal path based on its local environment. The operator is thus freed to direct more cognition to the analysis of sensor data represented in the control interface. Shared Mode allows

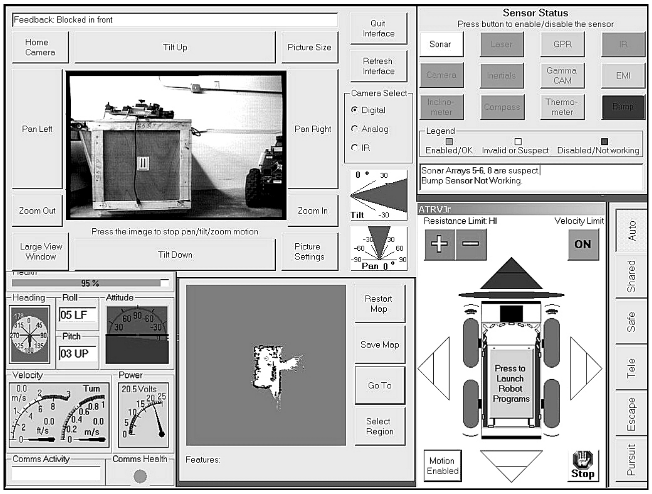

Figure 2. The status and control windows in the interface

the user to give directional input and, upon the identification of an item of interest, allows the operator to take control of the navigational tasks for further exploration of the area of interest. A final mode, *Autonomous Mode*, consists of a series of high-level tasks such as go to a point, patrol or search a selected region, or follow a designated path. In Autonomous Mode, the only user intervention occurs on the tasking level; the robot itself manages all robot navigation and obstacle avoidance.

Tele Mode most closely matches the current state of practice for remote system deployments. Safe Mode provides a level of initiative necessary to protect the robot and its environment from the missteps of novice system users. The previous study had indicated that the two most valuable modes were the Safe Mode and the Shared Mode. We wanted to further nuance the differences between these modes. Unfortunately, the first set of experiments masked the differences because, in a pattern consistent with a normal learning curve, the order in which each participant used the modes proved to be an overweening factor. In the present study, only Safe Mode and Shared Mode were used, and participants were only given one opportunity to run the robot. These two modes allowed a direct comparison between a "protected manual control" and "supervised navigational autonomy," respectively. We believe that these two modes represent the types of control scenarios that will be most relevant for use in a search and rescue scenario.

3.4 Procedure

The participants were informed they had 60 seconds to locate as many of the five items in the area (see Figure 3) as possible. Each participant was instructed on the use of the joystick for controlling the robot. Additionally each participant was instructed on the robot's camera controls (e.g. pan., tilt, zoom). For participants using Shared Mode, it was explained that they should let the robot control base-level navigation functions; however, if they wanted to redirect the robot, the robot would yield control to their joystick commands. In order to facilitate realistic maneuvering through an urban environment, the robot's search arena featured several obstacles. The central area was divided into quadrants using conventional office dividers, while the perimeter featured four pylons. Five objects were scattered throughout the arena and consisted of two dummies (representing injured humans), a stuffed dog, a disabled robot, and a small simulated explosive device. The participants controlled the robot from a remote station without direct line of sight, thereby ensuring maximum use of the available interface cues.

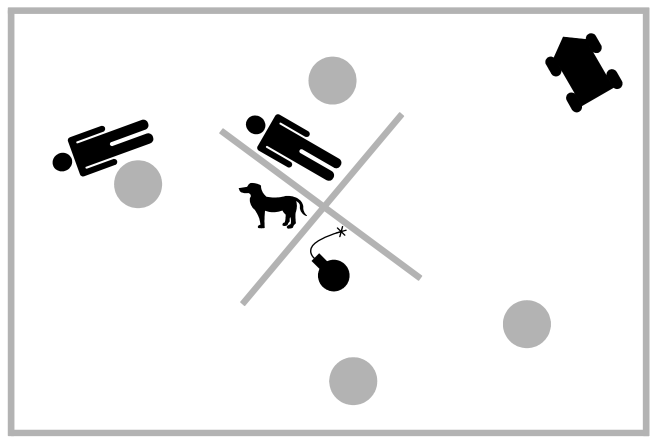

Figure 3. The search arena, featuring five objects (in black) and numerous obstacles (in grey)

3.5 Results

The effects of participant age, gender, and operational mode were compared against the total number of objects that were located and identified (see Figure 4). In contrast to our previous experiment [5] which emphasized subjective measures such as trust, ease-of-use, and feeling of control, this experiment focused on quantitative performance metrics. The participants were grouped by age in five-year intervals up to 20 years old; thereafter they were grouped in ten-year intervals. This ensured that the analysis was sensitive to possible developmental differences in pre-adults. There was no significant difference in the number of objects found across participants of different ages, $F(8,96)=1.64$, $p=0.12$. Note that the study did not feature a balanced sample across age groups. More young participants volunteered than did older adult participants. Although there were fluctuations in the number of objects that were found by different age

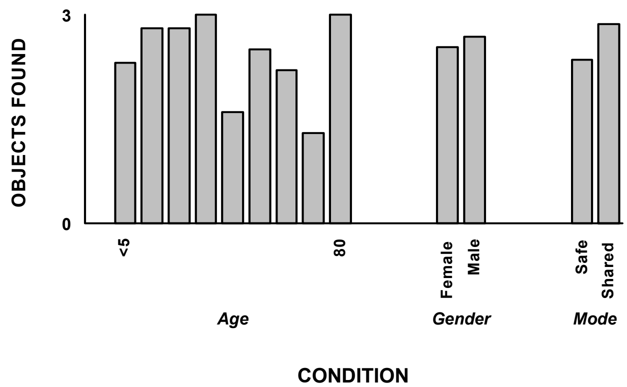

Figure 4. The average number of objects found (out of five possible) according to age, gender, and mode of operation

groups, the sample size was too small and there was insufficient statistical power to determine if those differences were meaningful.

The above analysis excludes two participants, who were eliminated from the analysis as statistical outliers. Two male participants were run back-to-back and were in the same age group (40-50 years old). Their performance at finding objects was more than two standard deviations above the average for other age groups. Because they were the only participants in the 40-50 year age group and because they exhibited unusually high performance compared to other participants, these two participants produced a significant statistical effect for their age group. It was concluded that their high performance was spurious. We would not expect their performance to reflect typical performance levels for their age group. Consequently, their data were not considered in the analyses.

There was no difference in the number of objects found due to gender. Females statistically found the same number of objects as did males, M=2.54 and M=2.68 respectively, $F(1,103)=0.31$, $p=0.58$.

There was a significant difference due to operational mode, $F(1,103)=4.83$, $p<0.05$. Participants who used Shared Mode found an average of 2.87 objects, while those who used Safe Mode found an average of 2.35 objects.

There were no significant twoway or threeway interactions between gender, age, and operational mode.

4 Discussion and Future Work

The results demonstrate that a very broad spectrum of novice users of a robotic interface were successfully able to operate the robot in an urban search and rescue scenario. Performance was significantly better in Shared Mode than in Safe Mode, suggesting that the system usability can be enhanced through the addition of navigational autonomy, freeing the user to focus on the search and rescue task instead of robot operation. In fact, although it is often assumed outside the robotics community that robot's can drive better than a human operator, this study provides some of the first compelling evidence that robot autonomy can actually out-perform human operators.

The results also highlight the value of iterative usability testing and redesign in making human-robot interfaces easy to use, as well as the value of carefully controlled HRI evaluation. Still, much work remains, and the present study is only a starting point for further HRI performance and usability evaluations.

In particular, an important area for ongoing research entails balancing performance and usability within the same HRI evaluation. Although our human-robot interface has undergone extensive usability testing elsewhere [5], the present study focused solely on the performance of users. User performance is an indirect measure of system usability. It would be useful to combine performance metrics with usability feedback from users during their operation of the robot. Future research will aim to integrate both performance and usability metrics into our HRI evaluations.

While the present approach to evaluating the human-robot interface capitalized on the performance of novice users, future work will balance findings from novice users and target or expert users. The types of performance issues common to novice users may not in all cases reflect the problems that can arise from the complex system interactions that the domain expert will experience.

Another important topic for future work concerns the role of navigational autonomy when multiple robots are controlled by a single operator or team. The complexity of manual teleoperation in a search and rescue task is multiplied when the operator must control multiple robots. Future work will investigate performance on a search and rescue task while controlling multiple robots in Safe Mode and Shared Mode. We hypothesize that the performance advantage of Shared Mode will be even more pronounced than it was in the current study.

5 Acknowledgement

This work is supported through the INEEL Long-Term Research Initiative Program under DOE Idaho Operations Office Contract DE-AC07-99ID13727.

6 References

[1] D. J. Bruemmer, J. L. Marble, and D. D. Dudenhoeffer, “Mutual initiative in human-machine teams,” *Proc. IEEE Conference on Human Factors and Power Plants*, pp. 7.22-30, 2002.

[2] D. J. Bruemmer, J. L. Marble, D. D. Dudenhoeffer, M. O. Anderson, and M. D. McKay, “Mixed-initiative control for remote characterization of hazardous environments,” *Proceedings of HICSS*, p. 127a, 2003.

[3] J. Casper, and R. Murphy, “Human-robot interactions during the robot assisted urban search and rescue response at the World Trade Center,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B,* Vol. 33, pp. 367-385, 2003.

[4] A. Jacoff, E. Messina, and J. Evans, “A reference test course for urban search and rescue robots,” *Proceedings of the SPIE-AeroSense Conference*, 2001.

[5] J. L. Marble, D. J. Bruemmer, and D. A. Few, “Lessons learned from usability tests with a collaborative cognitive workspace for human-robot teams,” *IEEE Conf. on Systems, Man, and Cybernetics*, 2003.

[6] D. R. Olsen, and M. A. Goodrich, “Metrics for evaluating human-robot interactions,” *Proceedings of PERMIS*, 2003.

[7] H. A. Yanco, and J. L. Drury, “A taxonomy for human-robot interaction,” *AAAI Technical Report FS-02-03*, pp. 111-119, 2002.

[8] H. A. Yanco, J. L. Drury, and J. Scholtz, “Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition,” *Journal of Human-Computer Interaction*, Vol. 19, pp. 117-149, 2004.